

Title	Spoken Dialogue System for Information Navigation based on Statistical Learning of Semantic and Dialogue Structure(Digest_要約)
Author(s)	Yoshino, Koichiro
Citation	Kyoto University (京都大学)
Issue Date	2014-09-24
URL	http://dx.doi.org/10.14989/doctor.k18614
Right	学位規則第9条第2項により要約公開; 許諾条件により本文は2016-06-01に公開
Type	Thesis or Dissertation
Textversion	ETD

(続紙 1)

京都大学	博士（情報学）	氏名	吉野 幸一郎
論文題目	Spoken Dialogue System for Information Navigation based on Statistical Learning of Semantic and Dialogue Structure 意味・対話構造の統計的学習に基づく情報案内のための音声対話システム		
(論文内容の要旨)			
<p>The thesis addresses a framework of a spoken dialogue system that navigates information of text documents, such as news articles, based on statistical learning of semantic and dialogue structure. Conventional spoken dialogue systems require a clear goal of the user, but this assumption does not always hold. Moreover, domain knowledge and task flows of the systems are conventionally hand-crafted. This process is costly and hampers domain portability.</p> <p>In this thesis, a task of information navigation that handles ambiguous user queries is proposed. The goal and procedure of the user are not well-defined, but the proposed system probes potential information demands of the user and presents relevant information proactively according to semantic similarity and dialogue context. The back-end domain knowledge of semantic structure is statistically learned from the domain corpus in an unsupervised manner. The proposed system has seven modules for information navigation. These modules are controlled using a dialogue manager based on statistical learning of dialogue structure.</p> <p>Chapter 2 introduces a scheme of information extraction, which is defined by the Predicate-Argument (P-A) structure and realized in unsupervised statistical learning. A useful information structure for information navigation depends on the domain, and the proposed significance score based on the Naive Bayes model selects a set of important P-A structures from the parsed results of domain texts. A preliminary experiment suggested that the significance score effectively extracts important patterns.</p> <p>Chapter 3 presents a novel text selection method for language modeling with Web texts for automatic speech recognition (ASR). Compared to the conventional approach based on the perplexity criterion, this method introduces a semantic-level relevance measure (significance score defined in Chapter 2) with the back-end knowledge base used in the dialogue system. It is used to filter semantically relevant sentences in the domain. Experimental evaluations in two different domains showed the effectiveness and generality of this method. The method combined with the perplexity measure results in significant improvement not only in ASR accuracy, but also in semantic and dialogue-level accuracy.</p> <p>Chapter 4 presents the spoken dialogue module that navigates not only exact information for the user query, but also partially-matched information related to user interests. Based on the information structure represented by the P-A structure, the dialogue module conducts question answering and proactive information presentation. In conjunction with the P-A significance score defined in Chapter 2, similarity measures of the P-A components are introduced to select relevant information. An experimental evaluation showed that the proposed system makes more relevant responses compared with the conventional system based on the ``bag-of-words'' scheme. The proactive presentation module was also implemented based on the relevance measure to the dialogue history.</p>			

Chapter 5 addresses an empirical approach to managing the proposed spoken dialogue system based on partially observable Markov decision process (POMDP). The POMDP has been widely used for dialogue management and is formulated for information navigation as a selection of modules. The POMDP-based dialogue manager receives a user intention and user focus, which are classified by spoken language understanding (SLU) based on discriminative models. These dialogue states are used for selecting appropriate modules by policy function, which is optimized by reinforcement learning. The reward function is defined by the quality of interaction to encourage long interaction of information navigation. Experimental evaluations with real dialogue sessions demonstrated that the proposed system outperforms the conventional rule-based system and the POMDP-based system that does not track the user focus in terms of dialogue state tracking and action selection.

Chapter 6 concludes this thesis. The proposed system is designed and trained in a domain-independent manner, so it can be ported to a variety of domains of the information navigation task.

注)論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し
審査結果の要旨は日本語500～2,000字程度で作成すること。

(論文審査の結果の要旨)

本論文は、意味と対話の構造の統計的な学習に基づいて、情報案内を行う音声対話システムに関する研究をまとめたものである。従来の音声対話システムが明確なタスクゴールを前提としていたのに対して、本システムではユーザの曖昧な情報要求に対応するとともに、ユーザの興味に応じて積極的な情報提示も行う点が特徴である。また、そのために必要なドメイン知識や対話制御を統計的に学習する方法について検討を行っている。具体的に得られた主な成果は以下の通りである。

1. 特定のドメインで情報案内を行う際に鍵となる意味構造を、述語項構造の重要度と類似度に基づいて定義し、これをコーパスから自動獲得する方法を提案した。これにより、ユーザの質問に対して適合する検索結果がなくても、関連する情報を提示することが可能になった。評価実験において、典型的な従来手法であるbag-of-wordsモデルと比べて、的確な応答生成の割合が増えることを示した。
2. 上記の述語項構造の重要度を用いて、Web上のテキストから音声認識の言語モデル学習に有用な文を選別する方法を提案した。これにより、当該ドメインに特化した音声認識システムを効率的に構成することが可能になった。評価実験において、従来の表層的な単語列の尤度に基づいて選別を行う手法と比べて、音声認識精度や意味理解精度を改善できることを示した。
3. 情報案内を行うための対話制御をPOMDP（部分観測マルコフ決定過程）で行う方法を提案した。これは、ユーザの焦点管理を導入した上で、情報案内における対話モジュールと報酬を定義し、強化学習を行うものである。評価実験において、従来のルールベースのシステムや焦点管理を行わないシステムと比べて、対話状態の推定精度や応答生成モジュール選択の成功率が改善されることを示した。

以上のように本論文は、情報案内という新たなカテゴリの音声対話システムの構成論を提示するもので、学術上・実用上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。また、平成26年8月21日に論文とそれに関連した内容に関する口頭試問を行った結果、合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。
更に、試問の結果の要旨（例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」）を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。
要旨公開可能日： 年 月 日以降